Opinion TRENDS in Genetics Vol.21 No.12 December 2005

On the origin of genomes and cells within inorganic compartments

Eugene V. Koonin[1] and William Martin[2]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA
[2]Institute of Botany III, University of Düsseldorf, Düsseldorf, Germany

Building on the model of Russell and Hall for the emergence of life at a warm submarine hydrothermal vent, we suggest that, within a hydrothermally formed system of contiguous iron-sulfide (FeS) compartments, populations of virus-like RNA molecules, which eventually encoded one or a few proteins each, became the agents of both variation and selection. The initial darwinian selection was for molecular self-replication. Combinatorial sorting of genetic elements among compartments would have resulted in preferred proliferation and selection of increasingly complex molecular ensembles – those compartment contents that achieved replication advantages. The last universal common ancestor (LUCA) we propose was not free-living but an inorganically housed assemblage of expressed and replicable genetic elements. The evolution of the enzymatic systems for (i) DNA replication; and (ii) membrane and cell wall biosynthesis, enabled independent escape of the first archaebacterial and eubacterial cells from their hydrothermal hatchery, within which the LUCA itself remained confined.

Introduction

Because all cells synthesize their proteins with the help of similar mechanisms on homologous ribosomes using the same genetic code, the same L-amino acids and the same D-sugars, biologists generally agree that all known life shares a common ancestor. However, the nature of that last universal common ancestor (LUCA) is a matter of conjecture and debate [1–13]. The traditional approach to inferring the nature of any common ancestor, including the LUCA, is to identify features that are shared by all of its descendants (or most of them, allowing for degenerative evolution), thereby pinning attributes to the ancestor. But that approach runs into problems with the LUCA because, beyond the code, the cores of the translation and transcription systems, and the homochirality of sugars and amino acids (i.e. the exclusive use of one stereoisomer of each, namely L-amino acids and D-sugars), few attributes are truly universal to all cells, as genome comparisons attest [14–16]. Today, all known cells, including those of highly reduced parasites, are surrounded by membranes and their genomes encode >500 proteins [12]. However, the onset of life could not have been a leap from disorganized chemical compounds to fully fledged cells. It must have been a stepwise process, albeit, perhaps, rapid in terms of geological time [17]. Along the path from inanimate chemicals to the first free-living cells (ignoring all dead-ends), molecular complexity inevitably increased, and each transitional stage must have had means of heritable variation in a setting that permitted selection [18]. Scenarios to account for these crucial early stages of the evolution life strive for an uninterrupted sequence of steps that are logically linked to one another [13].

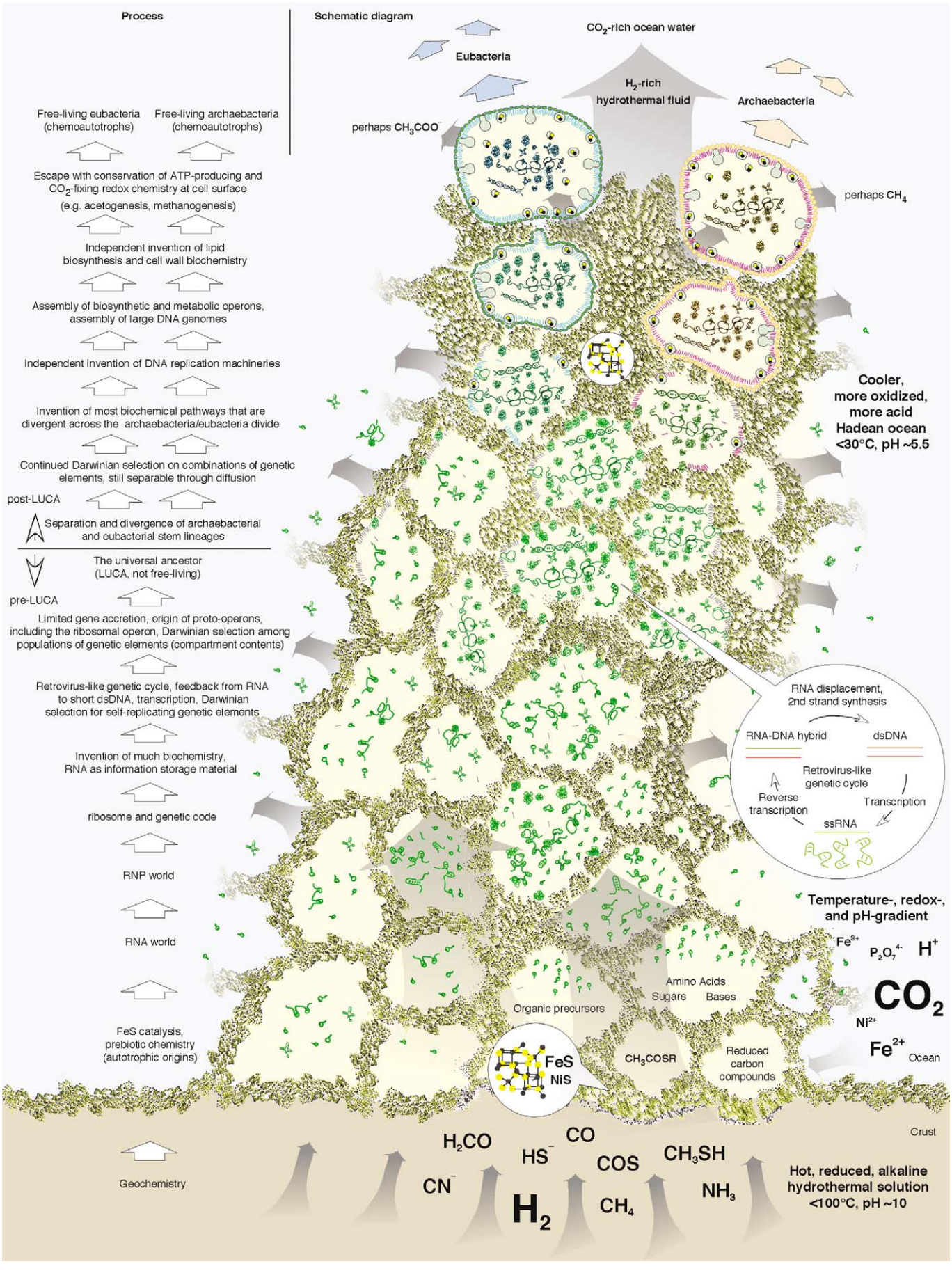
Modern biochemistry and information processing could have hardly evolved in free solution [19]. Without pre-existing and hence necessarily abiogenic compartments to restrain the building blocks of life from free diffusion into the ocean, the monomeric constituents of biopolymers would never have reached sufficient concentrations to react with one another [17,19,20]. Today, the barriers that organisms have against diffusion to the environment are lipid bilayer membranes and cell walls. However, archaebacteria (archaea) and eubacteria (bacteria), the two divisions of prokaryotes, have membranes consisting of unrelated lipids (isoprene ethers and fatty acid esters, respectively), and the enzymatic pathways involved in archaebacterial and eubacterial membrane biogenesis are non-homologous [21]; the cell walls of archaebacteria and eubacteria share even less chemical similarity [21,22].

The case for a LUCA confined within a network of inorganic compartments

Perhaps the simplest interpretation of the disparity of prokaryotic membranes is that the LUCA was not a free-living cell and did not have a biogenic membrane but, instead, arose and existed within geologically formed, inorganic, abiogenic compartments [17,19,20,23,24] that fulfilled the imperative compartmentalizing function of lipid bilayers and cell walls before the latter arose. Alternative views of the origin of cellular organization either suggest that life somehow emerged on two-dimensional surfaces and that cells bubbled off when the process had been completed [10,25] or merely postulate the transition from some form of uncontained prebiotic soup to cellular organization [6,26], without providing any plausible reasoning for how the complex chemistry and genetics of living systems could have arisen in free solution.

Corresponding author: Koonin, E.V. (koonin@ncbi.nlm.nih.gov).
Available online 11 October 2005

 0168-9525/$ - see front matter doi:10.1016/j.tig.2005.09.006


Process
Schematic diagram
Free-living eubacteria (chemoautotrophs)
Free-living archaebacteria (chemoautotrophs)
Escape with conservation of ATP-producing and CO_2-fixing redox chemistry at cell surface (e.g. acetogenesis, methanogenesis)
Independent invention of lipid biosynthesis and cell wall biochemistry
Assembly of biosynthetic and metabolic operons, assembly of large DNA genomes
Independent invention of DNA replication machineries
Invention of most biochemical pathways that are divergent across the archaebacteria/eubacteria divide
Continued Darwinian selection on combinations of genetic elements, still separable through diffusion
post-LUCA
Separation and divergence of archaebacterial and eubacterial stem lineages
The universal ancestor (LUCA, not free-living)
pre-LUCA
Limited gene accretion, origin of proto-operons, including the ribosomal operon, Darwinian selection among populations of genetic elements (compartment contents)
Retrovirus-like genetic cycle, feedback from RNA to short dsDNA, transcription, Darwinian selection for self-replicating genetic elements
Invention of much biochemistry, RNA as information storage material
ribosome and genetic code
RNP world
RNA world
FeS catalysis, prebiotic chemistry (autotrophic origins)
Geochemistry
CO_2-rich ocean water
Eubacteria
H_2-rich hydrothermal fluid
Archaebacteria
perhaps CH_3COO^-
perhaps CH_4
Cooler, more oxidized, more acid Hadean ocean <30°C, pH ~5.5
RNA displacement, 2nd strand synthesis
RNA-DNA hybrid
dsDNA
Retrovirus-like genetic cycle
Reverse transcription
Transcription
ssRNA
Temperature-, redox-, and pH-gradient
Fe^{3+}
$P_2O_7^{4-}$
H^+
CO_2
Ni^{2+}
Fe^{2+}
Ocean
Amino Acids
Sugars
Bases
Organic precursors
FeS
NiS
CH_3COSR
Reduced carbon compounds
Crust
H_2CO
HS^-
CO
COS
CH_3SH
CN^-
H_2
CH_4
NH_3
Hot, reduced, alkaline hydrothermal solution <100°C, pH ~10

Among the plethora of ideas about the origin of life, the hypothesis that living systems arose within naturally formed three-dimensional compartments is, in our view, currently the only one that can satisfy the demands imposed by the inordinate level of complexity that a biochemical and genetic system must attain before it can approximate life as we know it. From the geochemical standpoint, it seems most plausible that these compartments consisted of iron sulfide (FeS) deposited at a warm (<90 °C) submarine hydrothermal spring [17,24]. Such naturally formed, inorganic compartments ranging in size from <1 μm to ~100 μm, have been discovered at ancient hydrothermal sites [23], and their formation has been simulated in the laboratory [20,27,28], Moreover, intricate, small-scale networks of hydrothermally formed microcompartments (although not consisting of FeS owing to the paucity of Fe^{2+} in modern oceans) occur today at the warm ($\sim$40–90 °C) Lost City vent [29,30], which has been continuously active for $>30\,000$ years [31]. Iron sulfide compartments at a hydrothermal vent are conceptually satisfying 'hatcheries' for the emergence of life because they simultaneously provide a concentrating mechanism – physical restraint from free diffusion into the ocean – for biochemical building blocks, a catalyst for a wide variety of biochemical reactions in the form of Fe-S and Fe-Ni-S centers [17,32–34] and a continuous source of chemical energy (the H_2–CO_2 redox couple) that the origin of life inexorably requires [17,32] (Figure 1). In the present context, and consistent with prior formulations of hydrothermal origin of life [17,19,24], we demand only two reasonable attributes of the compartment system: (i) porosity, enabling movement of molecules of different sizes among compartments (with example in evidence at the active Lost City vents [30]); and (ii) sustained formation of new compartments through continuous metal-sulfide precipitation at the ocean interface (with example in evidence at more ancient vents [17]) such that organic molecules could invade new territories, thereby generating discrete units on which selection could act. Discrete units are essential, because in a homogeneous system selection is unattainable, whereas in a physically differentiated system it becomes possible [18].

According to the hydrothermal origin hypothesis, the first organisms were chemolithoautotrophs whose metabolism centered around thioester formation reactions [17, 19] similar to the modern acetyl-CoA (Wood-Ljungdahl) pathway [32]. These reactions would have been fuelled by the exergonic reaction between hydrothermal H_2 and marine CO_2 to produce reduced carbon compounds [32,35]. The model combines the plausibility of the FeS-based primordial biochemistry [34], in which carbonyl sulfide (COS; a catalyst of peptide bond formation [36,37]) might have played a crucial role, with a physical basis for compartmentalization rooted in geological observations [17,27]. Based on those considerations, we outline here how darwinian selection and evolution of genetic organization could have proceeded within the confines of such a warm hydrothermal vent. We stress 'warm' [24] (but not 'pond', as in Darwin's famous passage) because hot vents like the familiar 'black smokers' are typically 350–400°C [38], a temperature that is much too hot for the chemistry of life, at its origins or otherwise. Our proposal also draws on competition and selection among self-replicating molecules [39–42], the retrovirus-like model of the genetic system of LUCA [43,44] and the selfish operon hypothesis [45,46].

Replication of a confined LUCA

Archaebacteria and eubacteria sharply differ not only by virtue of their unrelated membranes and cell walls but also by the equally dramatic disparity of their DNA replication machineries [47,48]. Surprisingly, the central components of these systems (DNA polymerases, primases and replicative helicases) are either unrelated or, at least, not orthologous in archaebacteria and eubacteria [44,49–52]. This contrasts the basic unity and conservation among the major proteins of the translation and transcription systems [53]. Thus, LUCA is inferred to have had an advanced translation system that resembled modern ones in its principal features, but lacked a double-stranded, replicating DNA genome, possessing an RNA-based replication system instead [44,47,53].

However, several other components of the DNA replication machinery, such as the sliding clamp plus the clamp loader ATPase and the DNA ligase, as well as enzymes of DNA precursor biosynthesis – ribonucleotide reductase and thymidylate kinase – are homologous in all prokaryotes, which led to the proposal that LUCA had a retrovirus-like replication cycle [44] (Figure 1). Because RNA molecules are fragile compared with DNA, and no RNA virus with a monopartite genome >30 kb has been described [54], it has been suggested that the genome of LUCA consisted of a set of co-inherited RNA segments, each coding for one or a few proteins [44]. The retrovirus-like genetic cycle of LUCA would account for a set of multiplying, competing, functionally diversifying and recombining molecules without demanding the complexity of a fully fledged prokaryotic genome. For longer RNA molecules to unfold for replication, short-term temperature cycles (analogous to PCR reactions) would be extremely helpful in such a system. This does not require specially fluctuating environmental conditions because, in

Figure 1. A scenario of evolution from the origin of RNA molecules to the independent escape of archaebacterial and eubacterial cells within naturally forming inorganic compartments consisting primarily of FeS at a Hadean (>3.8 Ga old) hydrothermal vent. The left hand side of the figure shows the proposed sequence of events (from the bottom to the top) described in the main text; the right portion is a highly schematic drawing illustrating increasingly complex levels of molecular organization within the compartments along the path from inorganic carbon to LUCA and from LUCA to free-living chemoautotrophic prokaryotes (modified from Ref. [24]). The enlarged compartment shows the proposed retrovirus-like genetic cycle of LUCA [44]. The synthesis of the RNA–DNA hybrid and the dsDNA is thought to have been catalyzed by reverse transcriptase, with RNase H (a ubiquitous, probably, ancestral enzyme) involved in RNA strand removal during the latter step. The transcription of RNA from dsDNA would have been catalyzed by DNA-dependent RNA polymerase (another universal enzyme). The parts of the replication system that are shared by archaebacteria and eubacteria (sliding clamp, clamp loader ATPase and DNA ligase) might have been involved in the dsDNA synthesis and/or transcription steps. The figure is schematic, and not drawn to scale. The compartments could be from $<$1–100 μm in diameter, as in fossil [24] and modern [30] vents (see main text for details).

the presence of a temperature gradient, as exists at vent–ocean interfaces [17,30], laminar convection within compartments under isothermic conditions produces the required melting and annealing effect, an efficient process in the laboratory known as 'convective PCR' [55]. The utility of convective PCR at the origin of life has been noted [55], and it offers efficient replication in chambers of 10–500 μm thickness, in line with the size of modern [30] and ancient [17,27] vent compartments (<1–100 μm), thus underpinning the hydrothermal origin model.

Constrained within a primordial network of FeS compartments, virus-like genetic elements would have been, at this stage of evolution, the agents both of natural variation (mutation and recombination) and of natural selection, initially, for self-replication only. The life histories of these elements would have involved diffusion between compartments, colonization of newly formed compartments, unavoidable RNA recombination [56] and replication in those compartments where appropriate combinations of genetic entities and monomeric precursors had accumulated. Competition for substrates available within a compartment is analogous to the classical *in vitro* darwinian experiments of Mills *et al*. [39] and Biebricher *at al*. [57] and would necessarily have entailed the emergence of autocatalytic, self-sustaining ensembles of molecules (Figure 1). The latest mathematical modeling results indicate that the emergence of such autocatalytic networks could be more probable than is often assumed [42,58].

Selfish cooperatives

One class of winners in the struggle for replication would consist of truly selfish elements, non-coding parasites exploiting the resources provided by other elements, as modern viroids and virusoids do [59,60]. However, 'altruistic' elements, such as those coding for translation system components or products promoting nucleic-acid-precursor synthesis also would be selected for the ability to enhance their own replication potential, as well as that of others, through a 'cell-type strategy' [61]. Additional survivors would include elements coding for a RNA replicase and/or a nucleic-acid-binding protein. These can be seen as virus-like agents that depend on other elements for everything but replication, yet might promote replication of other genetic elements unable to do so on their own. In this regard, we note that an origin of viral-like genetic systems (although not fully fledged viruses) concomitant with the origin of the first genomes is a salient feature of the present model. At the first stages of evolution of replicating molecule ensembles, the viral-like and would-be prokaryotic genomes are indistinguishable in as much as selection operates solely at the level of self-replication. However, the formation of selfish cooperatives would mark the emergence of larger 'protogenomic' ensembles from viral-sized genetic systems and birth of viruses as genetic elements employing parasitic strategies; these elements could have had various replication cycles, some of which conceivably have persisted into the modern viral world, including RNA viruses, retroid viruses and, later, DNA viruses.

The sorting of individual genetic elements between compartments would result in new combinations leading to preferred proliferation of the genetic compartment contents that attained effective replication – selfish cooperatives. With increasing probability, the most successful compartments would infect their neighbors via diffusion of rapidly replicating combinations of genetic elements. Propagation of such ensembles would also involve molecular seeding of newly formed compartments at the continuously growing (enlarging through continuous FeS precipitation) surface of the vent. Replicative success would correspond to fitness in the classical sense, by virtue of seeding ability. Thus, the genetic complements of successful compartments would be the fittest, spreading such that these gene ensembles (fragmented 'genomes' of the compartments) would proliferate effectively, but with a much lower fidelity than modern genetic systems. Importantly, the genetic contents of the compartments would emerge as discrete agents of selection within a continuously expanding system of inorganically formed territories. Under this model, the geological formation of new compartments can be viewed as the abiogenic predecessor of cell division, whereas diffusion of coding molecules between compartments would be analogous to horizontal gene transfer (HGT) before the origin of free-living cells, as others have suggested previously, albeit in a context lacking physicochemical mechanisms to demark recipients from donors [6,62]. In contrast to earlier views, the present model operates via selection acting on discrete individual units (compartment contents) as is required for darwinian lineage differentiation [18] rather than via (essentially, unattainable) selection in a freely mixing gene pool.

In this scenario, there will be genetic parasites that avidly compete for resources. But just as today, parasites that eradicate their host eradicate themselves. Hence our selfish cooperators, although unable to outcompete their most aggressive parasites, could also potentially benefit from assimilating some of the greedy properties of the more attenuated parasites (e.g. high-fidelity replicases). Persistence of ensembles of selfish cooperators, particularly those with a weak autocatalytic feedback into resource synthesis [42,58], would be favored for long-term survival, albeit facing the constant threat of takeover by parasites. This aspect of the present model is similar to the switch from individual to group selection (selection for proto-cell gene packages) in the stochastic corrector model described by Szathmary [63–65].

Selfish superoperons

At this point, all routes to further complexity and sustenance of novelty would require co-distribution of a growing ensemble of genetic elements with a replicable translation system, a prototype of the ribosome, which LUCA unquestionably possessed. Originally, this proto-ribosome might have functioned as a triplet replicator (a triplicase), its function in translation arising later [13,66,67]; however, genome comparisons clearly indicate that LUCA already had a translation system similar in its basic features to the modern one [53].

The spread of advantageous combinations would be greatly facilitated if (some of) the genes for proteins that are, directly and indirectly, required for replication were located on the same genomic segment, along the lines of the selfish operon concept [45,46]. Multigenic elements roaming the LUCA network would be quintessential selfish operons, free of mobility constraints imposed by the large, typically, monopartite genomes of modern prokaryotes. The few operons that are shared by archaebacteria and eubacteria and do not show signs of extensive HGT could have descended from such primordial genetic elements. The primary winner in this primordial darwinian competition might have been an element encoding the principal ribosomal proteins and RNA polymerase subunits, a probable ancestor of the largest superoperon which is, to a varying extent, conserved in all contemporary archaebacteria and eubacteria [68,69].

Transition to the next complexity level would require longer, more stable information storage and transmission molecules, perhaps approaching large contemporary plasmids in size and coding volume, to counter the deleterious effects of reassortment disrupting beneficial combinations of genetic elements. But how big could such RNA genomes get? Poole and Logan [70] and, independently, Forterre [71] have recently suggested that LUCA could have had a highly complex RNA genome by virtue of RNA repair, which seems to be rather widely distributed among modern cell lineages, and which would improve the replication fidelity of an RNA genome. RNA-repair activities are congruent with the idea that LUCA, in principle, could have had a relatively large RNA genome. However, even with the benefit of several enzymes implicated in the repair of viral RNA [72], the largest known RNA genomes, those of coronaviruses, do not exceed ~30 kb [54], a size compatible with the formation of a complex superoperon but obviously insufficient to encode the entire repertoire of cellular functions on one or a few genomic segments. Ultimately, switching to DNA-based information storage and the accompanying functional specialization of DNA and RNA into the information-storing and information-expressing components of the genetic system, respectively, would have been a necessity *en route* to life as we know it.

DNA genomes and cellular escape

The synthesis of double stranded DNA (dsDNA) is an intrinsic feature of the retrovirus-like replication cycle postulated for the inorganically confined LUCA discussed above (Figure 1) and, obviously, would be inherited by the still inorganically confined ancestors of archaebacteria and eubacteria. However, it appears from genome comparisons that LUCA did not have the ability to replicate these small dsDNA molecules [44] and, because of this, the retrovirus-like cycle was not conducive to transmission of large numbers of linked genes. Genome comparisons suggest instead that the enzymatic systems for dsDNA replication, and hence the ability to use DNA as the main information storage device, evolved separately in the two primary prokaryotic lineages [44,70]. The first DNA-replication systems, possibly of virus-like nature [73], would have stored information in compartments that provided hospitable conditions. In such compartments, the DNA elements would have become sinks accreting other genetic elements, perhaps, including the ribosomal superoperon, via reverse transcription into growing DNA molecules. Agglomeration of the numerous proteins and cofactors required for the formation of lipids and semipermeable membranes (and the even larger number of subordinate proteins needed to provide amino acids, coenzymes and other metabolites efficiently) was probably unattainable before the emergence of large DNA genomes. Hence, we find it unlikely that cells with largely or solely RNA-based genetic storage were capable of existing outside a network of inorganic compartments, particularly because precise cell-division machinery is a prerequisite for successful escape.

The two distinct systems for membrane lipid biosynthesis might have evolved on small, mobile entities that colonized distinct populations of FeS compartments and joined the growing DNA genomes. Once membranes and turgor-resistant cell walls evolved, there must have been numerous escape attempts, most of them failing as the ocean was increasingly nutrient-poor and inhospitable with increasing distance from the vent. The successful, independent escape of proto-cells with archaebacterial and eubacterial membranes had to await the assembly, on a single (or a few) chromosome(s), of a collection of genes encoding the full complement of functions necessary and sufficient to support the free-living lifestyle.

We only considered the two prokaryotic domains of life and disregarded the eukaryotes. This is because, in contrast to suggestions that eukaryotes are direct descendants of the first cells [8,74], we find the available evidence to indicate that eukaryotes emerged much later, through symbiosis between fully fledged archaebacteria and eubacteria [75–77] (further specification of the possible partners involved being irrelevant here). Such a symbiosis could only have occurred after the escape of both forms of prokaryotic life from the primordial system of inorganic compartments.

Inorganic compartments versus membrane-bounded cells as the means for confining the LUCA

It has been repeatedly argued that the complex molecular composition inferred for LUCA could not have been attained without prior evolution of biogenic-membrane-bounded cells [18,26,71,78], mainly because (i) compartmentalization is a prerequisite for the evolution of any complex system; and (ii) certain key membrane-associated enzymes, such as the signal recognition particle (SRP) and the proton ATPase, are conserved in eubacteria and archaeabacteria. The model of a compartmentalized, but inorganically confined LUCA obviates the first problem. However, the second problem – the conservation of certain membrane-associated functions in all modern forms of life – is more challenging. The ubiquity of the SRP (with its notable RNA component) and the proton ATPase across genomes, together with the clear split between archaebacterial–eukaryotic and eubacterial versions, suggests that these complexes were present in LUCA as opposed to spread subsequently via HGT (notwithstanding some

HGT for the proton ATPase [79]). Because the SRP inserts proteins into hydrophobic layers and ATPase requires a hydrophobic layer to function, this would seem to imply the existence of membranes in LUCA, apparently in contradiction to our arguments above concerning the late and independent emergence of lipid biosynthetic pathways. The essential distinction to be made is between a 'hydrophobic layer' and a 'biogenic membrane'. The latter requires elaborate suites of lineage-specific enzymes (given the unrelated isoprene ether versus fatty acid ester chemistries of the membrane lipids in archaebacteria and eubacteria, respectively). The former could consist preferentially of C8–C12 aliphatic acids, which are expected to arise from H_2 and CO_2 geochemically in studies of thermodynamic equilibria under warm (~50–100 °C) hydrothermal vent conditions [35]. Furthermore, in experiments simulating vent conditions, *de novo* synthesis of aliphatic acids up to C5 [80], their elongation by up to three carbons atoms [81,82] and their condensation [81,82] through inorganic catalysis only have been reported. Thus, the SRP and proton ATPase could have operated within hydrophobic layers provided naturally at the compartment surfaces of a hydrothermal vent, without the demand for genetically encoded mechanisms of lipid synthesis.

Some biologists might object to this very specific aspect of our current formulation. However, it finds support from theory [35] and experiments [34,80–82], whereas the two current alternatives are more problematic. In one, Cavalier-Smith [10] argued that archaebacteria arose from actinobacteria (high-GC Gram-positive bacteria) ~850 million years ago (Mya) and evolved an entirely new lipid membrane and cell wall biochemistry in response to thermophilic adaptation; independently, Gupta proposed that archaebacteria derive from low-GC Gram-positive bacteria [83]. The problems we see with these scenarios are that: (i) no known prokaryotes have demonstrably undergone any vaguely similar cataclysmic lipid transition; and (ii) to our knowledge, no genome-wide data implicate either actinobacteria or low-GC Gram-positive bacteria as ancestors of archaebacteria. The other hypothesis, developed by Wächtershäuser [25], suggests that the LUCA was a form of life that existed in two dimensions only and that could synthesize both lipid (and, implicitly, cell wall) types, followed by differential loss. Differential loss explains all of the differences between archaebacteria and eubacteria, but makes the (primitive?) LUCA the biochemically most-potent organism that ever lived, with functionally redundant parallel pathways for a plethora of essential functions (lipids, cell walls, DNA replication). Our proposal avoids arguably the most challenging conundrum faced by all models entailing a free-living LUCA – replacement of one ancestral membrane type by another in either eubacteria or archaebacteria – and furthermore avoids the need for two-dimensional life, while keeping the LUCA metabolically simple and simultaneously accounting for the observations from genomes.

We cannot summarily dismiss models entailing a free-living LUCA, but the model of an abiogenically confined LUCA outlined here seems to require less imagination with respect to membrane evolution, and is also satisfying in other respects. In particular, it explicitly permits evolution of considerable complexity without requiring prior emergence of a complete cell division mechanism (such a mechanism, even a primitive one, is prerequisite to the free-living state) or mechanisms for intercellular DNA exchange that are required for HGT, and are central to some competing views. On the contrary, this model allows locally restrained transfer of coding material within a system of physically juxtaposed, interconnected compartments – a suitable setting for the emergence of genetic complexity.

Concluding remarks

This formulation of hydrothermal origins builds on previous models [17,19,20,23,24,27] and specifically addresses the emergence and early evolution of genetic systems. It does not tackle the origin of the genetic code, although it seems to be compatible with the recent, intriguing suggestion that the code evolved through a two-letter intermediate entailing the RNA-catalyzed synthesis of biogenic amino acids on RNA [84]. The model does, however, offer conceptual inroads towards structuring severe problems – limiting steps – *en route* to evolving genomes. What would keep genes from diffusing into the ocean? What was the source of energy supporting their synthesis? How could the transition from selection for self-replication to selection for versatile functions have occurred? A continuously expanding set of naturally formed compartments (with real examples in nature) would enable early (geo)biochemistry and self-replicating systems to attain a level of complexity approaching that of the simplest extant life forms. The transition from non-living to living systems in pre-existing compartments that have a continuous energy and carbon source, in which combinations of self-replicating molecular cooperatives can arise and undergo *bona fide* darwinian selection, implies a LUCA that was complex but dramatically different from modern cells, most notably in that it was not free-living. We believe that this version of LUCA – physically contained, possessing ribosomes and translation but lacking membranes and large DNA genomes – could be a crucial and not implausible intermediate in the early evolutionary sequence.

How can these ideas be tested? More work needs to be performed with chemistry simulating hydrothermal vent conditions (reviewed in Ref. [34]), in particular, to investigate synthesis of cofactors and bases in the presence of suitable nitrogenous compounds [24] (the synthesis of peptides via carbonyl sulfide seems to work efficiently [36,37]). If microbiology and genomics were to uncover a dense spectrum of continuous variation spanning the archaebacterial–eubacterial divide, such that their deep distinctness should disappear (e.g. prokaryotic groups with archaebacterial DNA replication but eubacterial membranes or *vice versa*), the abiogenically confined LUCA aspect would become unnecessary. Our prediction is clearly that, despite much HGT between archaebacteria and eubacteria, genuine intermediates will never be found, because they never existed as free-living cells.

Acknowledgements

We thank M.J. Russell and Y.I. Wolf for many helpful discussions and for critical reading of the article, P. Forterre and D. Penny for providing preprints of their work, and the Intramural Research Program of the NIH (E.V.K.) and the DFG (W.M.) for funding.

References

1 Darnell, J.E. and Doolittle, W.F. (1986) Speculations on the early course of evolution. *Proc. Natl. Acad. Sci. U. S. A.* 83, 1271–1275
2 Forterre, P. *et al*. (1992) The nature of the last universal ancestor and the root of the tree of life, still open questions. *BioSystems* 28, 15–32
3 Koch, A.L. (1994) Development and diversification of the Last Universal Ancestor. *J. Theor. Biol.* 168, 269–280
4 Doolittle, W.F. and Brown, J.R. (1994) Tempo, mode, the progenote, and the universal root. *Proc. Natl. Acad. Sci. U. S. A.* 91, 6721–6728
5 Woese, C. (1998) The universal ancestor. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6854–6859
6 Woese, C.R. (2002) On the evolution of cells. *Proc. Natl. Acad. Sci. U. S. A.* 99, 8742–8747
7 Lazcano, A. and Forterre, P. (1999) The molecular search for the last common ancestor. *J. Mol. Evol.* 49, 411–412
8 Penny, D. and Poole, A. (1999) The nature of the last universal common ancestor. *Curr. Opin. Genet. Dev.* 9, 672–677
9 Doolittle, R.F. (2000) Searching for the common ancestor. *Res. Microbiol.* 151, 85–89
10 Cavalier-Smith, T. (2002) The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* 52, 7–76
11 de Duve, C. (2003) A research proposal on the origin of life. *Orig. Life Evol. Biosph.* 33, 559–574
12 Koonin, E.V. (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* 1, 127–136
13 Penny, D. An interpretative review of the origin of life research. *Biol. Philos.* (in press)
14 Koonin, E.V. (2000) How many genes can make a cell: the minimal-gene-set concept. *Annu. Rev. Genomics Hum. Genet.* 1, 99–116
15 Harris, J.K. *et al*. (2003) The genetic core of the universal ancestor. *Genome Res.* 13, 407–412
16 Charlebois, R.L. and Doolittle, W.F. (2004) Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res.* 14, 2469–2477
17 Russell, M.J. and Hall, A.J. (1997) The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc. (London)* 154, 377–402
18 de Duve, C. (2005) The onset of selection. *Nature* 433, 581–582
19 Russell, M.J. *et al*. (1994) A hydrothermally precipitated catalytic iron sulphide membrane as a first step toward life. *J. Mol. Evol.* 39, 231–243
20 Russell, M.J. *et al*. (1989) *In vitro* growth of iron sulphide chimneys: possible culture chambers for origin-of-life experiments. *Terra Nova* 1, 238–241
21 Boucher, Y. *et al*. (2004) Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol. Microbiol.* 52, 515–527
22 Kandler, O. and Konig, H. (1998) Cell wall polymers in Archaea (Archaebacteria). *Cell. Mol. Life Sci.* 54, 305–308
23 Russell, M.J. *et al*. (1988) Submarine hot springs and the origin of life. *Nature* 336, 117
24 Martin, W. and Russell, M.J. (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 59–83
25 Wächtershäuser, G. (2003) From pre-cells to Eukarya – a tale of two lipids. *Mol. Microbiol.* 47, 13–22
26 Pereto, J. *et al*. (2004) Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem. Sci.* 29, 469–477
27 Russell, M.J. *et al*. (2003) On the dissipation of thermal and chemical energies on the early Earth: The onsets of hydrothermal convection, chemiosmosis, genetically regulated metabolism and oxygenic photosynthesis. In *Natural and Laboratory-Simulated Thermal Geochemical Processes* (Ikan, R., ed.), pp. 325–338, Kluwer Academic Publishers
28 Stone, D.A. and Goldstein, R.E. (2004) Tubular precipitation and redox gradients on a bubbling template. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11537–11541
29 Kelley, D.S. *et al*. (2001) An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30 degrees N. *Nature* 412, 145–149
30 Kelley, D.S. *et al*. (2005) A serpentinite-hosted ecosystem: the Lost City hydrothermal field. *Science* 307, 1428–1434
31 Fruh-Green, G.L. *et al*. (2003) 30,000 years of hydrothermal activity at the lost city vent field. *Science* 301, 495–498
32 Russell, M.J. and Martin, W. (2004) The rocky roots of the acetyl-CoA pathway. *Trends Biochem. Sci.* 29, 358–363
33 Svetlitchnyi, V. *et al*. (2004) A functional Ni-Ni-[4Fe-4S] cluster in the monomeric acetyl-CoA synthase from Carboxydothermus hydrogenoformans. *Proc. Natl. Acad. Sci. U. S. A.* 101, 446–451
34 Cody, G.D. (2004) Transition metal sulfides and the origins of metabolism. *Annu. Rev. Earth Planet. Sci.* 32, 569–599
35 Shock, E.L. *et al*. (1998) The emergence of metabolism from within hydrothermal systems. In *Thermophiles: The Keys to Molecular Evolution and the Origin of Life* (Wiegel, J. and Adams, M.W.W., eds), pp. 59–76, Taylor and Francis
36 Huber, C. and Wachtershauser, G. (1998) Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: implications for the origin of life. *Science* 281, 670–672
37 Leman, L. *et al*. (2004) Carbonyl sulfide-mediated prebiotic formation of peptides. *Science* 306, 283–286
38 Douville, E. *et al*. (2002) The Rainbow vent fluids (36°14′ N, MAR): the influence of ultramafic rocks and phase separation on trace metal content in Mid-Atlantic Ridge hydrothermal fluids. *Chem. Geol.* 184, 37–48
39 Mills, D.R. *et al*. (1967) An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. U. S. A.* 58, 217–224
40 Eigen, M. *et al*. (1980) Hypercycles and compartments. Compartments assists–but do not replace–hypercyclic organization of early genetic information. *J. Theor. Biol.* 85, 407–411
41 Eigen, M. and Schuster, P. (1977) The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* 64, 541–565
42 Hordijk, W. and Steel, M. (2004) Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. Theor. Biol.* 227, 451–461
43 Wintersberger, U. and Wintersberger, E. (1987) Retroviruses and the origin of life. *Trends Genet.* 3, 198–202
44 Leipe, D.D. *et al*. (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res.* 27, 3389–3401
45 Lawrence, J. (1999) Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr. Opin. Genet. Dev.* 9, 642–648
46 Lawrence, J.G. and Roth, J.R. (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143, 1843–1860
47 Mushegian, A.R. and Koonin, E.V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10268–10273
48 Edgell, D.R. and Doolittle, W.F. (1997) Archaea and the origin(s) of DNA replication proteins. *Cell* 89, 995–998
49 Leipe, D.D. *et al*. (2000) The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome Res.* 10, 5–16
50 Keck, J.L. *et al*. (2000) Structure of the RNA polymerase domain of *E. coli* primase. *Science* 287, 2482–2486
51 Tye, B.K. (2000) Insights into DNA replication from the third domain of life. *Proc. Natl. Acad. Sci. U. S. A.* 97, 2399–2401
52 Forterre, P. (2002) The origin of DNA genomes and DNA replication proteins. *Curr. Opin. Microbiol.* 5, 525–532
53 Anantharaman, V. *et al*. (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* 30, 1427–1464
54 Gorbalenya, A.E. (2001) Big nidovirus genome. When count and order of domains matter. *Adv. Exp. Med. Biol.* 494, 1–17
55 Brown, D. and Goddard, N.L. (2003) Exponential DNA replication by laminar convection. *Phys. Rev. Lett.* 91, 158103-158101-158103-158104

56 Gmyl, A.P. *et al*. (2003) Nonreplicative homologous RNA recombination: promiscuous joining of RNA pieces? *RNA* 9, 1221–1231
57 Biebricher, C.K. *et al*. (1985) Kinetics of RNA replication: competition and selection among self-replicating RNA species. *Biochemistry* 24, 6550–6560
58 Mossel, E. and Steel, M. (2005) Random biochemical networks: the probability of self-sustaining autocatalysis. *J. Theor. Biol.* 233, 327–336
59 Symons, R.H. (1991) The intriguing viroids and virusoids: what is their information content and how did they evolve? *Mol. Plant Microbe Interact.* 4, 111–121
60 Diener, T.O. (2001) The viroid: biological oddity or evolutionary fossil? *Adv. Virus Res.* 57, 137–184
61 Nemoto, N. and Husimi, Y. (1995) A model of the virus-type strategy in the early stage of encoded molecular evolution. *J. Theor. Biol.* 176, 67–77
62 Kandler, O. (1994) The early diversification of life. In *Early Life on Earth* (Bengston, S., ed.), pp. 152–160, Columbia Univ. Press
63 Szathmary, E. and Demeter, L. (1987) Group selection of early replicators and the origin of life. *J. Theor. Biol.* 128, 463–486
64 Zintzaras, E. *et al*. (2002) Living under the challenge of information decay: the stochastic corrector model vs. hypercycles. *J. Theor. Biol.* 217, 167–181
65 Brosius, J. (2003) Gene duplication and other evolutionary strategies: from the RNA world to the future. *J. Struct. Funct. Genomics* 3, 1–17
66 Altstein, A.D. (1987) Origin of the genetic system. *Mol. Biol.* 21, 309–322
67 Poole, A. *et al*. (1999) Early evolution: prokaryotes, the new kids on the block. *BioEssays* 21, 880–889
68 Lathe, W.C., 3rd. *et al*. (2000) Gene context conservation of a higher order than operons. *Trends Biochem. Sci.* 25, 474–479
69 Wolf, Y.I. *et al*. (2001) Genome alignment, evolution of prokaryotic genome organization and prediction of gene function using genomic context. *Genome Res.* 11, 356–372
70 Poole, A.M. and Logan, D.T. (2005) Modern mRNA proofreading and repair: clues that the Last Universal Common Ancestor (LUCA) possessed an RNA genome? *Mol. Biol. Evol.* 22, 1444–1455
71 Forterre, P. (2005) The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie* 87, 793–803
72 Snijder, E.J. *et al*. (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331, 991–1004
73 Forterre, P. (1999) Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Mol. Microbiol.* 33, 457–465
74 Brinkmann, H. and Philippe, H. (1999) Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Mol. Biol. Evol.* 16, 817–825
75 Martin, W. and Muller, M. (1998) The hydrogen hypothesis for the first eukaryote. *Nature* 392, 37–41
76 Lopez-Garcia, P. and Moreira, D. (1999) Metabolic symbiosis at the origin of eukaryotes. *Trends Biochem. Sci.* 24, 88–93
77 Rivera, M.C. and Lake, J.A. (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431, 152–155
78 Jeffares, D.C. *et al*. (1998) Relics from the RNA world. *J. Mol. Evol.* 46, 18–36
79 Hilario, E. and Gogarten, J.P. (1993) Horizontal transfer of ATPase genes–the tree of life becomes a net of life. *Biosystems* 31, 111–119
80 Heinen, W. and Lauwers, A.M. (1996) Organic sulfur compounds resulting from the interaction of iron sulfide, hydrogen sulfide and carbon dioxide in an anaerobic aqueous environment. *Orig. Life Evol. Biosph.* 26, 131–150
81 Cody, G.D. *et al*. (2000) Primordial carbonylated iron-sulfur compounds and the synthesis of pyruvate. *Science* 289, 1337–1340
82 Cody, G.D. *et al*. (2004) Assaying the catalytic potential of transition metal sulfides for abiotic carbon fixation. *Geochim. Cosmochim. Acta* 68, 2185–2196
83 Gupta, R.S. (1998) Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* 62, 1435–1491
84 Copley, S.D. *et al*. (2005) A mechanism for the association of amino acids with their codons and the origin of the genetic code. *Proc. Natl. Acad. Sci. U. S. A.* 102, 4442–4447